

SYSTEM AND METHOD FOR ROUTER DATA AGGREGATION AND DELIVERY

Tony M. Brewer
5225 Mariners Drive
Plano, Texas 75093
Citizenship: U.S.A.

Harry C. Blackmon
2517 LaVida Pl.
Plano, Texas 75023
Citizenship: U.S.A.

Chris Davies
8755 Southwestern #2122
Dallas, Texas 75206
Citizenship: U.S.A.

Harold W. Dozier
6906 McKamy Blvd.
Dallas, Texas 75248
Citizenship: U.S.A.

Thomas C. McDermott III
265 Daniel Dr.
Plano, Texas 75094
Citizenship: U.S.A.

Steven J. Wallach
7314 Wester Way
Dallas, Texas 75248
Citizenship: U.S.A.

Dean E. Walker
1829 Chattham Ct.
Plano, Texas 75025
Citizenship: U.S.A.

Lou Yeh
1432 Harrington Drive
Plano, Texas 75075
Citizenship: U.S.A.

RELATED APPLICATIONS

Subs
a1

This application is related to concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P001US-10020638], entitled "System And Method For IP Router With an Optical Core," to concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P002US-10020639], entitled "System and Method for Router Central Arbitration," to concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P006US-10020643], entitled "Timing and Synchronization for an IP Router Using an Optical Switch," to concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P012US-10021641], entitled "Router Network Protection Using Multiple Facility Interfaces," to concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P013US-10021642], entitled "Router Line Card Protection Using One-for-N Redundancy" and to concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P014US-10021643], entitled "Router Switch Fabric Protection Using Forward Error Correction," the disclosures of which are incorporated herein by reference.

TECHNICAL FIELD

This application relates to the field of optical communication networks, and particularly to large-scale routers for optical communication networks.

BACKGROUND

In an optical router the information that passes through the core switching fabric is a chunk of information that must go from the input side of a line shelf through the optical switch itself to the egress side of a switch shelf. No storage is available in the optical switch itself. Therefore the information must be formatted in such a way that data passes through the optical switch and is collected on the egress side, such that every chunk that is collected can potentially require a different phase adjustment for clock and data recovery. Further, since the chunk size is significantly larger than the minimum size packet that is brought into the system, there needs to be a more efficient method to transfer small packets through the optical switch itself.

An optical system exhibits cross-talk and relatively high loss characteristics. Thus, errors that are incurred through the optical switch require error checking and correcting for a high degree of reliability. To pick a given chunk out of a stream of random bits requires framing information. Additionally, it is required to have a reasonably equal mix of zeroes and ones with reasonable limitation of continuous streams of zeroes and ones for DC balance going through the optical switch, resulting in DC balance of light coming into the optical receiving circuitry from the switch itself.

SUMMARY OF THE INVENTION

The present invention is directed to a chunk format, which solves the various problems to allow a purely optical switch to operate at the core of a large-scale, high data throughput router. The format includes a preamble that allows each individual chunk to have clock and data recovery performed before the actual information within the chunk is received.

5 The format further includes a chunk header that contains information specific to the entire chunk. The chunk according to the present format can contain multiple packets, with each packet having its own packet header for packet-specific information. The format provides for a scrambler seed which allows scrambling the data to achieve a favorable zero and one balance as well as minimal run lengths. There are many multiple scrambler seed values that are available to be used, such that there can be a random choice of scrambler seed for any particular chunk to avoid the malicious forcing of zero and one patterns or run length of bits zeroes and ones. There are a chunk cyclical redundancy check (CRC) as well as forward error correction (FEC) bytes to detect and/or correct any errors and also to insure a high degree of data and control integrity. Advantageously, a framing symbol inserted into the

10 chunk format itself allows the receiving circuitry to identify or locate a particular chunk format. "Break Bytes" and "Make Bytes" fields located at the beginning of a chunk preamble precondition an optical receiver to a proper state before the actual chunk arrives at the receiver. A "Break Bytes" field is configured to maintain a 50% density of 0's and 1's for the transmit laser during the switch fabric "dark period." The receiver threshold level is

15 quickly reestablished by setting the value of a "Make Bytes" field to all ones. The "Make Bytes" are followed by preamble bytes, which are alternating ones and zeroes.

20

Subs
a 2
25

Various aspects of the invention are described in concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P001US-10020638], entitled "System And Method For IP Router With an Optical Core," concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P002US-10020639], entitled "System and Method for Router Central Arbitration," concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P006US-10020643],

Subs
A2

entitled "Timing and Synchronization for an IP Router Using an Optical Switch," concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P012US-10021641], entitled "Router Network Protection Using Multiple Facility Interfaces," concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P013US-10021642], entitled "Router Line Card Protection Using One-for-N Redundancy" and concurrently filed, co-pending, and commonly assigned U.S. Application Serial Number [59182-P014US-10021643], entitled "Router Switch Fabric Protection Using Forward Error Correction," the disclosures of which are incorporated herein by reference.

The foregoing has outlined rather broadly the features and technical advantages of the present invention in order that the detailed description of the invention that follows may be better understood. Additional features and advantages of the invention will be described hereinafter which form the subject of the claims of the invention. It should be appreciated by those skilled in the art that the conception and specific embodiment disclosed may be readily utilized as a basis for modifying or designing other structures for carrying out the same purposes of the present invention. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims. The novel features which are believed to be characteristic of the invention, both as to its organization and method of operation, together with further objects and advantages will be better understood from the following description when considered in connection with the accompanying figures. It is to be expressly understood, however, that each of the figures is provided for the purpose of illustration and description only and is not intended as a definition of the limits of the present invention.

BRIEF DESCRIPTION OF THE DRAWING

For a more complete understanding of the present invention, reference is now made to the following descriptions taken in conjunction with the accompanying drawing, in which:

Fig. 1 is a block diagram illustrating the format of a data chunk in a communication network router; and

Fig. 2 is a block diagram illustrating the format of the chunk header of Fig. 1 in more detail.

5

DETAILED DESCRIPTION

A communication network router is a system that distributes information from multiple sources to multiple destinations in the network. At the core of a router is a switching fabric that directs the information from the multiple sources to the multiple destinations. A fabric is generally defined as a collection of devices that cooperatively perform routing functions. In accordance with embodiments of the present invention, all information flows through the fabric in substantially fixed sized data chunks. In some embodiments the switch fabric contains an optical switch.

Fig. 1 is a block diagram illustrating the format of a data chunk 10 in a communication network router. The chunk format starts with a preamble 101-1 programmable in length, sufficient to allow router clock and data recovery circuitry to properly phase align the circuitry to retrieve data out of the chunk. Two fields "Break Bytes" 101-3 and "Make Bytes" 101-2 respectively are located at the beginning of chunk preamble 101-1. The purpose of these fields is to precondition an optical receiver to a proper state before the actual chunk arrives at the receiver. The optical switch when changing configuration passes very little light through to the output fibers. During this period of time, designated the "dark period," the receivers receive no light and thus need to reestablish the decision threshold level of the limiting amplifier within the burst mode optical receiver. The threshold level is quickly reestablished by setting the value of the "Make Bytes" field 101-2 to all ones. "Make Bytes" 101-2 are followed by preamble bytes 101-1, which are alternating ones and zeroes.

The "Break Bytes" field 101-3 is configured to maintain a 50% density of 0's and 1's for the transmit laser. "Break Bytes" field 101-3 will have a higher density of zeros in order to balance the all ones in "Make Bytes" field 101-2. The "Break Bytes," "Make Bytes," and preamble fields are all programmable in length to allow flexibility. It is noted that the "Break Bytes" and "Make Bytes" fields can alternatively be located at the trailing end of the string of chunk fields in Fig. 1, since chunks are transmitted one after the other, end to end. Thus the "Break Bytes" and "Make Bytes" could correctly be placed before the preamble or after the

trailing end of a chunk.

Referring again to Fig. 1, a chunk header 102 contains information including chunk type that is common to the entire chunk. There are four types of chunks, namely idle, administrative, guaranteed bandwidth and best effort chunks. Idle chunks are used strictly to travel between ASICs within the system as placeholders to maintain framing alignment. Administrative chunks are used to perform administrative functions across the optical switch. Administrative functions include dynamic alignment of chunks that are sent through the optical switch as well as keep-alive messages from a source to a destination of the optical switch to insure that the path is usable. Guaranteed bandwidth chunks carry guaranteed bandwidth traffic that is placed in those chunks, and best effort chunks carry best effort traffic. Best effort traffic is further self-divided into three best effort quality of service (QOS) levels, namely, best effort 1, 2 and 3. Guaranteed bandwidth chunks are separated from best effort chunks to allow guaranteed bandwidth traffic to pass ahead of best effort traffic through the optical switch, whereas best effort traffic stays in its original order through the switch.

Fig. 2 is a block diagram illustrating the format of chunk header 102 in more detail and showing a chunk type field 102-1. Additionally, within chunk header 102 is a specific routing subplane 102-2 through which the chunk will pass. In an embodiment of the present invention, there are six subplanes within the system. Accordingly, the routing subplane specifies through which of those six subplanes an individual chunk is destined. The routing subplane is used on the source side of the switch to steer the chunk to that specified subplane and on the destination side to check that the chunk was received on the specified subplane. This verifies the connectivity of the optical fibers connecting the subplanes to the receivers. Chunk header 102 includes a header parity 102-3 to verify that the chunk header is valid, as well as source and destination IDs 102-4 and 102-5 respectively to specify the source port and the destination port and to verify again that the fiber connectivity is correct.

Chunk header 102 includes a master chunk bit 102-6, which specifies that a particular chunk is a master chunk, such that every sixteenth chunk sent through the switch is designated as a master chunk. Chunk header 102 also contains a sequence number 102-7,

providing for separate sequence streams for guaranteed bandwidth chunks and for best effort chunks. The sequence number is unique within a source/destination pair, such that a destination is able to look at the last sequence number received for a particular source and guarantee that no chunks were corrupted beyond the point of being able to use them. The sequence number is used for alarming and for alerting that a chunk potentially was corrupted. Further, in chunk header 102 there is a re-initialize bit 102-8 indicating that sequence number 102-7 should be re-initialized at the destination. This is used to enable sequence number re-initialization to avoid the alarming mechanism.

Each individual packet within chunk 10 has its own header, for example, fields 103-1 and 103-2, which includes information specific to that packet. Packet header information specifies whether a packet segment contained in the chunk is a packet start or a packet end, if that packet segment is the entire packet including both start and end, or whether the packet segment is the middle of the packet. This information is used for reassembling multi-chunk packets as well as for specifying whether a packet is contained completely within the chunk. Additionally, contained in a packet header is a byte count specifying the number of bytes contained in the packet segment associated with this particular packet header. Also included is a bit, which if active, indicates that the packet should never be discarded. This bit is frequently set for a guaranteed bandwidth packet in chunks marked as guaranteed bandwidth chunks. Best effort 1, 2 and 3 classes should be designated only if the chunk has been indicated as a best effort chunk. There is a Point-to-Point Protocol (PPP) header format specifying how the destination facility module should treat the packet in terms of what PPP format should be appended to the chunk as it is being sent out. Packet header 103-1, 103-2 also contains a bit indicating whether the packet should be sent out through the output of the router or whether it should be looped back into the destination packet forwarding engine to be used by that packet forwarding engine. Packet header 103-1, 103-2 also includes a destination tributary indicator specifying to which tributary at a destination port the packet should be sent.

Fields 104-1 and 104-2 within the chunk format are the actual payloads of the packets

associated with respective packet headers 103-1 and 103-2. Packet header/payload pairs, for example 103-1 and 104-1, can be contained within the chunk payload up to a limit on the order of nine of these pairs, due to the 400 byte total payload size of a chunk versus a minimum packet size of 40 bytes. A chunk filler 105 is the next field. If there are not sufficient data in packets to fill the chunk payload, then the unused payload capacity is filled with non-traffic bytes.

The next field is a scrambler seed 106. The scrambler seed value is used to insure an equal balance of zeroes and ones across a chunk as well as to minimize the run lengths of ones and zeroes respectively across the chunk. A particular scrambler algorithm used for the chunk format does not insure a particular run length of zeroes and ones but minimizes that run length by scrambling of the bits.

The next field is a two-byte chunk CRC 107 used to insure that the overall chunk consisting of fields 102 through 107 has the proper CRC value once it is corrected by the forward error correcting (FEC) information described below in more detail. Fields 102 through 107 in the present router implementation contain a total of 400 bytes, including four bytes for chunk header 102, four bytes for each packet header 103-1, 103-2, a flexible number of bytes for packet payloads, for example 104-1, 104-2, two bytes for scrambler seed 106, two bytes for CRC 107, and the remainder in chunk filler 105.

The next field in sequence is a framing symbol 108, which is 16 bytes in length and is used to determine where in a stream of bits beginning and trailing end of a chunk are located. Framing symbol 108 is a unique symbol that must be found within a window roughly 10 nanoseconds in size, in order to identify a chunk. The framing symbol could be found anywhere within those 10 nanoseconds and it must be guaranteed that the framing symbol is unique within those 10 nanoseconds worth of bits.

Sixteen-byte framing symbol 108 is located near the trailing end of the chunk format, because the trailing end of the chunk format is a more reliable place than the beginning of the chunk to position the framing symbol. The beginning of the chunk format near preamble 101-1 has a potential for being unreliable, in that it takes more than the expected number of

bits to acquire the phase of the incoming signal of bits of information, and the first few bits of the fields at the beginning of the chunk may become corrupted whereas the trailing end of the chunk should not suffer from that problem. Accordingly, framing symbol 108 is located near the trailing end of chunk 10, to insure with the highest probability that it is received without errors. Since framing symbol 108 is not covered by FEC field 109, the framing symbol must be received without any correction applied to it. For greater reliability two bytes of 16-byte framing symbol 108 are placed just before FEC field 109, and 14 bytes worth of bits are intermixed into the preceding 56 bytes of fields. Accordingly each of these preceding 56 bytes contains two bits of framing symbol 108 and six bits of actual data.

In an embodiment of the present invention, a 400-byte chunk payload is delivered to an internal optics ASIC, which generates 32 FEC bytes. Once the 32 FEC bytes are generated, then 16-byte framing symbol 108 is inserted into the chunk payload. Viewing the 400-byte chunk payload as a bit stream, framing symbol 108 is inserted as two bits of framing symbol per six bits of data at the trailing end of the 400-byte chunk payload. Of the 16-byte framing symbol, 2 bytes are inserted after the 400-byte payload, and the other 14 bytes are intermixed two bits at a time per 6 bits of other data. The two bits that are intermixed always have the same value, a zero and a one bit. These two bits ensure that, when the internal optics ASIC scans the payload within the framing window, it will find no other framing symbol pattern until it actually finds the intended framing symbol. Once the framing symbol bits are inserted, then the 32 FEC bytes are appended. Once the framing symbol is inserted, then the intermixed fields are no longer byte aligned fields.

In operation, a receiving ASIC looks for the framing symbol. There is enough variation in arrival times of chunks that the receiving ASIC starts looking for the framing symbol up to 56 bytes before the actual framing symbol. To make sure that the receiving ASIC does not incorrectly find another string of bits having the same pattern as the framing symbol, 14 bytes of the framing symbol are forced two bits per byte into the preceding 56 bytes of the chunk. The receiving ASIC finds the framing symbol and then stops looking. Accordingly there is no need to force in extra bits after the framing symbol. If the receiving

ASIC is still looking for the framing symbol after the point where it should be located, it means that the framing symbol is corrupted. If the framing symbol is corrupted, then the receiving ASIC relies on the redundancy in the system.

Although the present invention and its advantages have been described in detail, it should be understood that various changes, substitutions and alterations can be made herein without departing from the spirit and scope of the invention as defined by the appended claims. Moreover, the scope of the present application is not intended to be limited to the particular embodiments of the process, machine, manufacture, composition of matter, means, methods and steps described in the specification. As one of ordinary skill in the art will readily appreciate from the disclosure of the present invention, processes, machines, manufacture, compositions of matter, means, methods, or steps, presently existing or later to be developed that perform substantially the same function or achieve substantially the same result as the corresponding embodiments described herein may be utilized according to the present invention. Accordingly, the appended claims are intended to include within their scope such processes, machines, manufacture, compositions of matter, means, methods, or steps.